Supplemental Figures

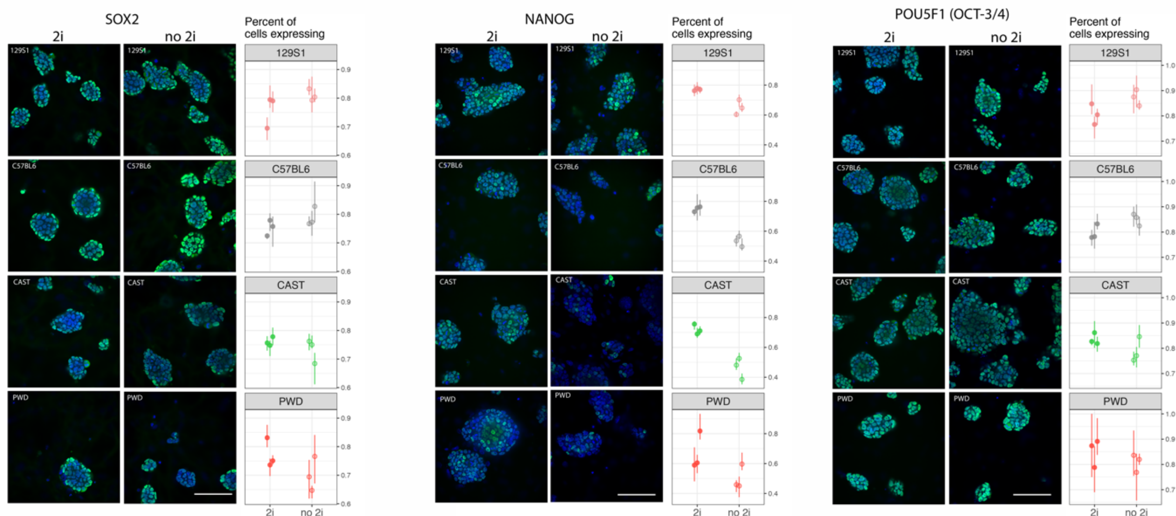

Figure S1, related to Figure 1. Immunofluorescence and quantitative analysis of microscopy images in four diverse mESC backgrounds. Quantification of SOX2, NANOG, and OCT4 expression in mESC lines constructed from diverse inbred backgrounds using immunofluorescence. For each protein representative images of mESCs from each genetic background grown with and without 2i are shown. To the right of representative images, dot plots show data collected from the same lines but analyzed quantitatively using high-content imaging and analysis software (Methods). Each dot represents an estimate of the percentage of cells expressing the protein for one replicate of the mESC line, summed across nine randomly selected images from different regions of the well. Error bars show the range in estimation of the percentage of cells expressing the protein across five images in the Z stack.

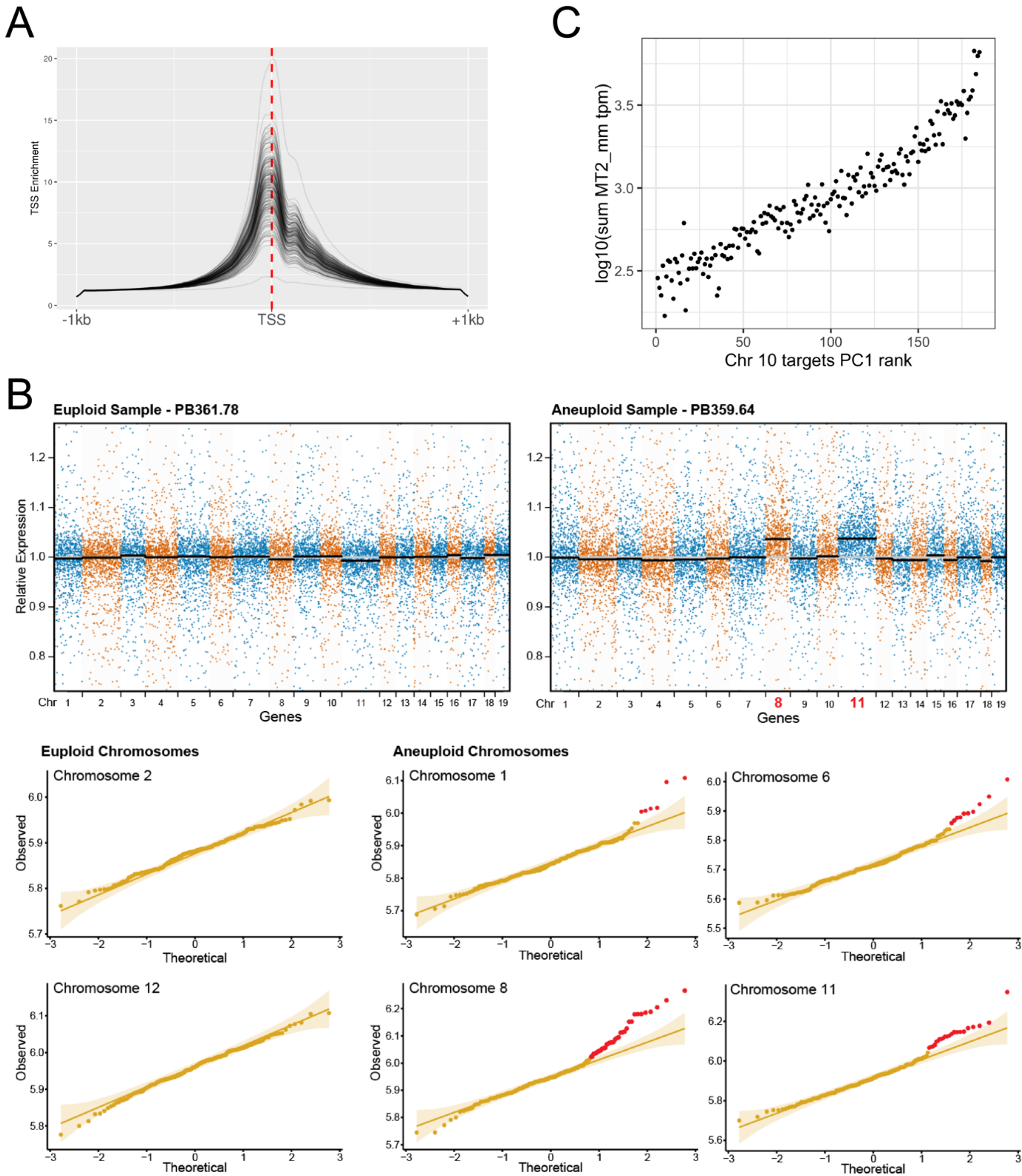

Figure S2, related to Figure 2. Features of DO mESC genetic elements, chromatin accessibility, and transcription. (A) Open chromatin at the transcription start site (TSS) across 16,956 genes annotated by Gencode (Frankish et al., 2019). ATAC-Seq read depth was calculated at each TSS +/- 1kb, depth was normalized to one at the -1kb and +1kb boundaries, and relative depth is plotted. Each line shows a separate DO mESC sample. (B) Gene expression data can be used to identify DO lines that have acquired chromosomal duplications. Within a sample, each gene's expression relative to the population median is

plotted on the *y*-axis against the gene's chromosome on the *x*-axis (top two plots). Within a chromosome, genes are ordered by position (but plotted as equidistant to more clearly display trends). ESC lines that do not have an appreciable proportion of aneuploid cells (top left) show the median gene expression of each chromosome nearing the population median (black bars are the sample chromosome medians, grey line is the population median = 1). ESC lines with larger proportions of aneuploid cells (top right) deviate from the population median for the duplicated chromosomes (Chromosomes 8 and 11 at bottom). Bottom six plots show the distribution of sample chromosome medians, which can reveal ESC lines with duplications of a given chromosome. For chromosomes euploid across all samples, we assume that the sample median expression values will fit a normal distribution (representative euploid chromosomes shown at bottom left). Chromosomes that are aneuploid in ESC lines should violate this assumption, with aneuploid samples exhibiting higher median expression values that fall outside the normal distribution (bottom right). Using these criteria, chromosomes 1, 6, 8, and 11 were classified as aneuploid in multiple DO lines (Shapiro-Wilk Normality Test $p > 0.05$). (C) Correlation between transcription of the MERVL long terminal repeat MT2_mm (http://www.dfam.org/entry/DF0004155) and Chr 10 hotspot QTL target genes. Reads from each DO mESC sample were aligned to a transcriptome consisting of all annotated genes (Ensembl release 82) as well all non-redundant mouse genome hits to the Dfam (Hubley et al., 2016) HMM for this repeat element using kallisto (Bray et al., 2016) version 0.43.1. Figure shows samples ranked by principal component 1 of the expression matrix of Chr 10 hotspot QTL targets, plotted against the summed transcripts per million for all MT2_mm transcripts. Individual points represent individual DO mESC samples.

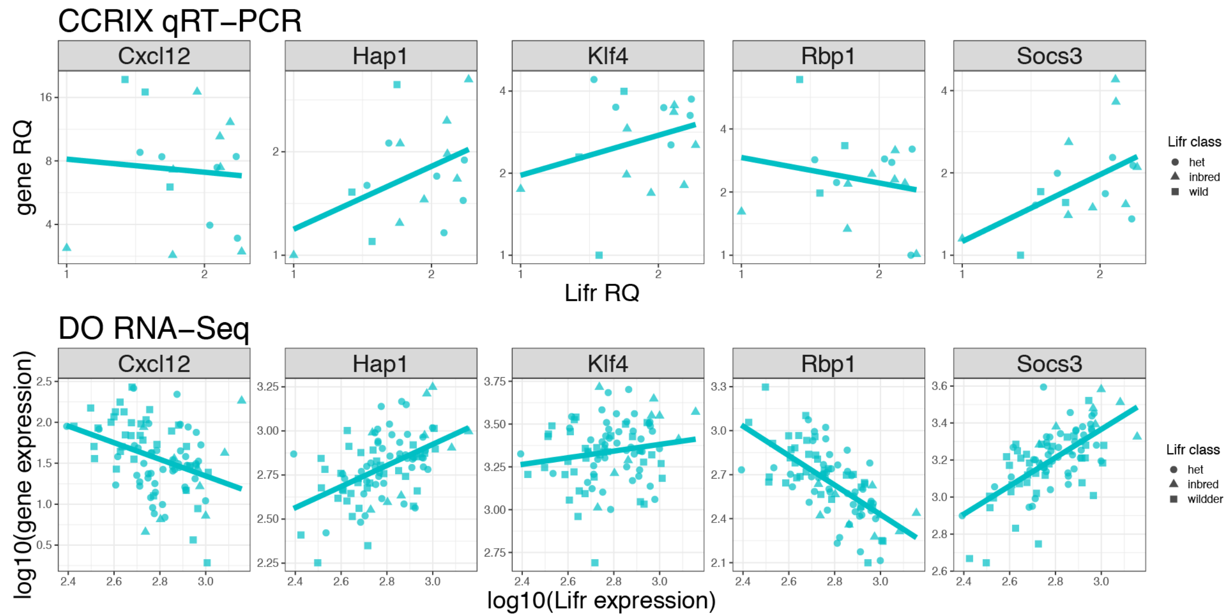

Figure S3, related to Figure 3. Quantitative PCR validation of five gene expression hotspot targets. qRT-PCR validation of five genes with eQTL mapping to the Chr 15 hotspot. We compared RQ (relative quantification via qPCR) of *Lifr* to each of the other five genes (facets). Top panels show qPCR of CC-RIX mESCs, while bottom panels show gene expression inferred from bulk RNA-Seq of DO mESCs. Only male CC-RIX mESC lines were used, thus data from DO mESC lines is restricted to only male lines. Note that gene co-expression relationships show similar trends (lines show best-fit regression) despite different genetic resource populations and methods of quantitation.

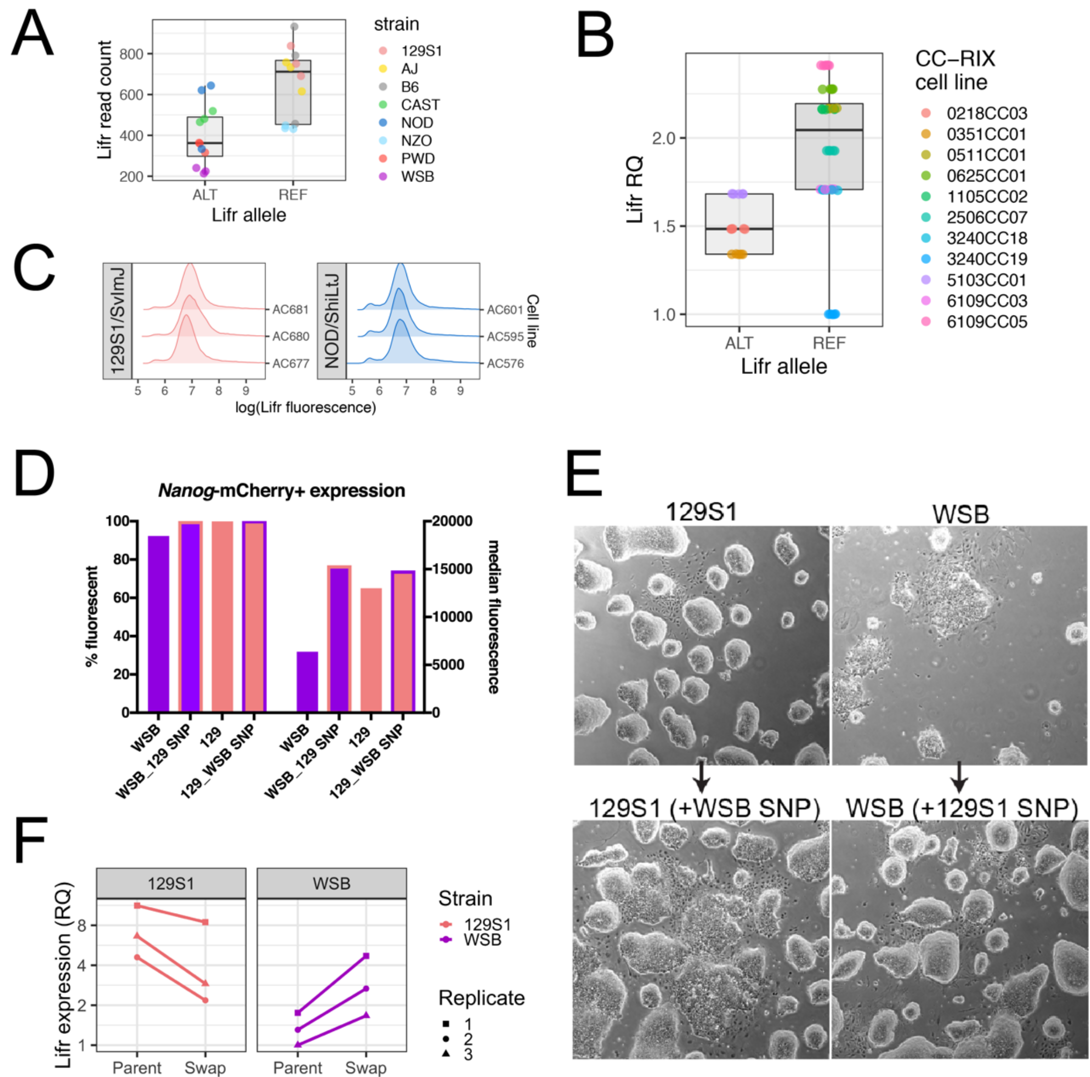

Figure S4, related to Figure 4. Validating genetically driven variation in *Lifr* expression and the effect of allele swaps on NANOG expression. (A) *Lifr* genotype affects *Lifr* expression as measured by RNA-Seq in CC founder strains. Points plotted represent individual data values. In each summary boxplot the horizontal lines show first quartile, median, and third quartile, while the whiskers extend a maximum of 1.5 times the inter-quartile range. (B) *Lifr* genotype affects *Lifr* expression as measured by qPCR in CC-RIX mESCs (only homozygous lines are shown). RQ indicates relative quantification via qRT-PCR. Boxplots are as in (A). (C) Flow cytometry using an antibody to LIFR provides quantitative assessment of the relative intensity of LIFR protein expression where NOD mESCs consistently have a sub-population of LIFR low cells. (D) Flow cytometry provides quantitative assessment of the percentage of cells expressing and relative intensity of *Nanog*-mCherry expression. Figure shows *Nanog*-mCherry expression in representative 129S1 and WSB parental mESC lines, as well as

the allele swap clones. In 2i culture conditions, the WSB parental mCherry lines have a slightly reduced percentage of cells expressing mCherry, but more significantly reduced median fluorescence. The positive effect of the allele swap in WSB cells with regard to the percentage of cells expressing and median fluorescence is apparent, however we do not observe the opposing effect of the allele swap in 129S1. (E) Colony morphology of WSB mESC clones harboring the classic inbred strain allele improves under 2i growth conditions. The opposing effect was somewhat less consistent in 129S1 mESC clones harboring the wild-derived-like allele. (F) Under 2i growth conditions, *Lifr* expression is higher in WSB mESC clones that harbor the REF allele (T/T; “swap”) with respect to the parental WSB mESC, and lower in 129S1 mESC clones that harbor the ALT allele (A/A; swap) with respect to the parental 129S1 mESC line. Quantitative RT-PCR data for three replicate experiments are shown.